

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2005-128922

(43)Date of publication of application : 19.05.2005

(51)Int.Cl.

G06F 13/00

H04L 12/58

(21)Application number : 2003-365887

(71)Applicant : NEC SOFTWARE KYUSHU LTD

(22)Date of filing : 27.10.2003

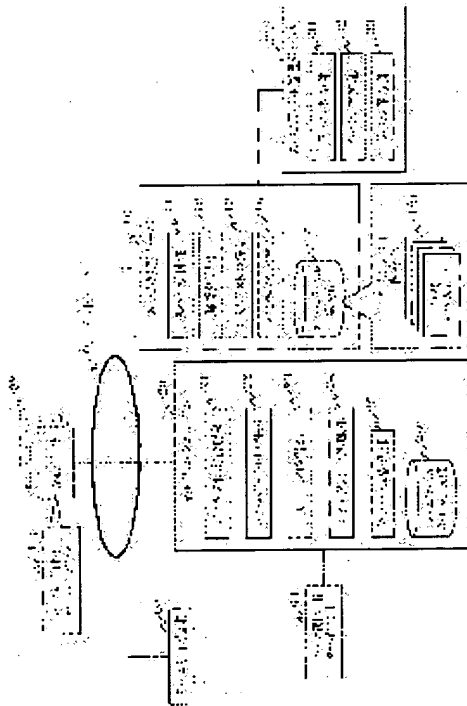
(72)Inventor : UEDA KENJI

(54) SPAM MAIL FILTERING SYSTEM AND METHOD AND ITS PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a spam mail filtering system capable of holding a URI showing the resource of a Web site included in a spam mail, and filtering a spam mail based on the held URL at the time of referring the resource.

SOLUTION: This spam mail filtering system is provided with: a Web site where an information resource including a Web page exists; a spam issuing terminal which issues a spam mail by referring to a mail address for accepting a spam registered in an arbitrary Web site in the Web site; a spam processing server which accepts the spam mail, and stores it as a URI spam list; a spam label management server which conforms URI included in the spam mail, and corrects any error; a client for resource browsing which allows a user to perform access to an information resource; and a filtering engine which executes the filtering of browsing based on a status preliminarily set by the URI spam list from the spam processing server at the time of executing the resource browsing of URI from the client for resource browsing.



【特許請求の範囲】**【請求項1】**

Webページを含む情報リソースが存在するWebサイトと、前記Webサイトの中の任意のWebサイトに登録されたスパム受付用メールアドレスを参照しスパムメールを発行するスパム発行端末と、前記スパムメールを受け付けてURIスパムリストとして保持するスパム処理サーバと、前記スパムメールに含まれるURIを取得して確認し誤りがあれば訂正するスパムラベル管理サーバと、情報リソースにアクセスするためのリソース閲覧用クライアントと、前記リソース閲覧用クライアントからのURIのリソース閲覧の際に前記スパム処理サーバからのURIスパムリストによって、あらかじめ設定した状態に基づいて閲覧をフィルタリングするフィルタリングエンジンと、前記Webサイト、前記スパム発行端末、前記スパム処理サーバ、および前記フィルタリングエンジンを相互に接続するネットワークとを有することを特徴とするスパムメールフィルタリングシステム。

【請求項2】

前記スパム処理サーバは、スパムメールを受け付けるスパム受付部と、前記スパムメールの内容を解析する解析処理部と、前記スパムメール内のURIをURIスパムリストとして登録するスパムリスト登録部と、前記URIスパムリストを保存するスパムラベル保存部と、前記スパムラベル保存部のURIスパムリストの問合せによって応答を行うスパムラベル応答部とを具備することを特徴とする請求項1記載のスパムメールフィルタリングシステム。

【請求項3】

前記スパムラベル管理サーバは、前記スパム処理サーバからURIスパムリストを取得するラベル取得部と、URIのリソースを取得して表示し当該リソースが有害か否かを確定するラベル確定部と、確定したURIを前記スパム処理サーバに送付して前記スパムラベル保存部の更新を指示するラベル保存部とを具備することを特徴とする請求項1または2記載のスパムメールフィルタリングシステム。

【請求項4】

前記フィルタリングエンジンは、前記URIスパムリストによってフィルタリングを行うか否かを含めてフィルタリング条件を設定するフィルタ設定管理部と、前記リソース閲覧用クライアントからのURI参照要求の際に前記スパム処理サーバからURIスパムリストを取得するスパムラベル取得部と、前記フィルタ設定管理部が設定したフィルタリング条件と前記スパムラベル取得部が取得したURIスパムリストとによってフィルタリングを行うフィルタリング処理部と、フィルタリング不要あるいは閲覧可能な場合に当該URIのリソースを取得しそれを前記リソース閲覧用クライアントに送付するリソース取得部とを具備することを特徴とする請求項1、2、または3記載のスパムメールフィルタリングシステム。

【請求項5】

前記フィルタリングエンジンは、前記スパムラベル取得部が取得したURIスパムリストをキャッシュとして保存するスパムラベルキャッシュ部およびそれを制御するキャッシュ管理部を備え、前記リソース閲覧用クライアントからのURI参照要求の際に当該URIを既に取得しているか否か前記スパムラベルキャッシュ部をチェックし、取得済みの場合にはそれを参照することを特徴とする請求項4記載のスパムメールフィルタリングシステム。

【請求項6】

前記スパム処理サーバは、あらかじめキーワードを登録しているキーワード保存部と、前記解析処理部がスパムメールを解析する際に前記スパムメールに含まれるキーワードを抽出し、それを前記キーワード保存部のキーワードと比較して当該キーワードの有害度を判断し、前記スパムメールを格付けするスパムラベル格付部と、前記スパムラベル格付部の格付結果に基づいて当該URIおよびそのスパムラベルをスパムラベル保存部に登録するスパムラベル登録部とを備えることを特徴とする請求項2記載のスパムメールフィルタ

リングシステム。

【請求項7】

前記Webサイトにスパムメールを受信可能な複数のメールアドレスを有しそれぞれのメールアドレスに同一のスパムメールを受信し、前記複数のメールアドレスの一つが受信したスパムメールを受付けてそれを前記スパム処理サーバへ送付する第二スパム処理サーバを備え、前記スパム処理サーバでは自身が受付けたスパムメールと前記第二スパム処理サーバから送付されたスパムメールとが同一のスパムメールであることを認識して解析処理することを特徴とする請求項1または2記載のスパムメールフィルタリングシステム。

【請求項8】

前記スパム処理サーバは、前記複数のメールアドレスの一つが受信した第一のスパムメールと前記複数のメールアドレスの他の一つが受信した第二のスパムメールとを比較し両者が同一のスパムメールであることを認識する際、それぞれのメールの内容を突き合せて照合し同じ内容を含む照合率に基づいて判定することを特徴とする請求項1、2、または7記載のスパムメールフィルタリングシステム。

【請求項9】

スパムメールを受け付け、前記スパムメール内に含まれるURIを収集しそれをURIスパムリストとして保持し、フィルタリングを行う際には、前記URIスパムリストを参照し、あらかじめ設定されているフィルタリング条件に基づいてフィルタリングすることを特徴とするスパムメールフィルタリング方法。

【請求項10】

スパムメールを解析する際に前記スパムメールに含まれるキーワードを抽出し、それをあらかじめ登録してあるキーワードと比較して当該キーワードの有害度を判断し、それをスパムラベルとして前記スパムメールを格付けし、その格付結果に基づいて当該URIおよびそのスパムラベルを保持し、フィルタリングを行う際には、前記URIおよびそのスパムラベルを参照し、あらかじめ設定されているフィルタリング条件に基づいてフィルタリングすることを特徴とするスパムメールフィルタリング方法。

【請求項11】

複数のメールアドレスの一つが受信した第一のスパムメールと前記複数のメールアドレスの他の一つが受信した第二のスパムメールとを比較し両者が同一のスパムメールであることを認識する際、それぞれのメールの内容を突き合せて照合し同じ内容を含む照合率に基づいて判定し、その判定結果を当該URIスパムリストに反映させることを特徴とする請求項9記載のスパムメールフィルタリング方法。

【請求項12】

スパムメールを利用して情報のフィルタリングを行う際に、スパムメールを受け付けるスパム受付機能と、前記スパムメールの内容を解析する解析処理機能と、前記スパムメール内のURIをURIスパムリストとして登録するスパムリスト登録機能と、前記URIスパムリストを保存するスパムラベル保存機能と、前記URIスパムリストへの問合せによって応答を行うスパムラベル応答機能とを、コンピュータに実現させるためのスパムメールフィルタリングプログラム。

【請求項13】

前記URIスパムリストによってフィルタリングを行うか否かを含めてフィルタリング条件を設定するフィルタ設定管理機能と、リソース閲覧用クライアントからのURI参照要求の際に前記URIスパムリストを取得するスパムラベル取得機能と、前記フィルタ設定管理機能が設定したフィルタリング条件と前記スパムラベル取得機能が取得したURIスパムリストとによってフィルタリングを行うフィルタリング処理機能と、フィルタリング不要あるいは閲覧可能な場合に当該URIのリソースを取得しそれを前記リソース閲覧用クライアントに送付するリソース取得機能とを、コンピュータに実現させるための請求項12記載のスパムメールフィルタリングプログラム。

【請求項14】

スパムメールを解析する際に前記スパムメールに含まれるキーワードを抽出し、それを

あらかじめ登録してあるキーワードと比較して当該キーワードの有害度を判断し、それをスパムラベルとして前記スパムメールを格付けし、その格付結果に基づいて当該URIおよびそのスパムラベルを保持し、フィルタリングを行う際には、前記URIおよびそのスパムラベルを参照し、あらかじめ設定されているフィルタリング条件に基づいてフィルタリングする機能を、コンピュータに実現させるためのスパムメールフィルタリングプログラム

【発明の詳細な説明】

【技術分野】

【0001】

本発明はスパムメールフィルタリングシステム、方法、およびプログラムに関し、特にスパムメールを利用して情報のフィルタリングを制御するスパムメールフィルタリングシステム、方法、およびプログラムに関する。

【背景技術】

【0002】

従来、有害なWebサイトのリソース（スパムまたはスパムメールという。）閲覧をフィルタリングする手法としては、インターネット上の情報にメタデータラベルを付与して、これをもとにフィルタリングする手法がある。メタデータラベルの付与には、情報発信者自らが付与するセルフレイティングの手法と、第三者が付与する手法がある。

【0003】

あるいはこのほか、情報に含まれる有害なキーワードや画像のイメージマッチングなどの照合処理によって、メタデータラベルを用いずにフィルタリングする手法もある。

【0004】

しかし、セルフレイティングの手法は、情報発信者が付与する手間や付与しても利用されない、あるいは不特定多数に閲覧させるために意図的に付与しない、などの理由から、普及していない。

【0005】

さらに、第三者レイティングの手法では、世界中の不特定多数の人が作成する情報は爆発的に増大していることから、新規に作成された情報やURI（uniform resource identifierの略。）が変更された情報にはメタデータラベルの付与が間に合わなかったり、あるいは膨大な費用を投じてメタデータラベルを付与する必要がある。あるいは、格付けには時間がかかることから、スパムメールが出回った直後にその情報が閲覧されてしまったあと、格付け結果が反映されるという状態になっており、期間限定の違法な情報を掲載したようなWebサイトのURIを含むスパムメールに対しては有効なフィルタリング手段ではない。

【0006】

キーワードやイメージマッチングなどの照合方式では、有効な情報まで誤ってフィルタリングされたり、あるいは有効な情報の検索さえできないといったマイナス面が発生することもある。

【0007】

上記のように、スパムメールに対する有効なフィルタリング手法は見出されていないが、最近ではスパムの発信源を発見した場合にはそれが発行する全電子メールの配信サービスをブロックするなどが提案されている（たとえば、特許文献1参照。）

【特許文献1】特表2001-527257号公報（第9～13頁，図1）

【発明の開示】

【発明が解決しようとする課題】

【0008】

上述のように、従来の情報フィルタリング方法では、個別情報に対する適切な格付けをタイムリーに取得し、それをフィルタリングに反映させることは困難である。また、発信源データに基づいて電子メールの配信を制御することにより、スパムメールなどの受信をブロックさせる手法もあるが、これは情報の内容を評価してフィルタリングしていることにはならない。すなわち、貴重な情報がスパムメールとして配信されない恐れもある。

【0009】

本発明の目的は、上記のような欠点を改善するために、スパムメールに含まれるWebサイトのリソースを指すURIを保持しておき、利用者がこのURIのリソースを参照する場合に、保持しているURIに基づいてフィルタリングするスパムメールフィルタリングシステム、方法、およびプログラムを提供することにある。

【課題を解決するための手段】

【0010】

本発明のスパムメールフィルタリングシステムは、Webページを含む情報リソースが存在するWebサイトと、前記Webサイトの中の任意のWebサイトに登録されたスパム受付用メールアドレスを参照しスパムメールを発行するスパム発行端末と、前記スパムメールを受け付けてURIスパムリストとして保持するスパム処理サーバと、前記スパムメールに含まれるURIを取得して確認し誤りがあれば訂正するスパムラベル管理サーバと、情報リソースにアクセスするためのリソース閲覧用クライアントと、前記リソース閲覧用クライアントからのURIのリソース閲覧の際に前記スパム処理サーバからのURIスパムリストによって、あらかじめ設定した状態に基づいて閲覧をフィルタリングするフィルタリングエンジンと、前記Webサイト、前記スパム発行端末、前記スパム処理サーバ、および前記フィルタリングエンジンを相互に接続するネットワークとを有することを特徴とする。

【0011】

さらに、本発明において、前記スパム処理サーバは、スパムメールを受け付けるスパム受付部と、前記スパムメールの内容を解析する解析処理部と、前記スパムメール内のURIをURIスパムリストとして登録するスパムリスト登録部と、前記URIスパムリストを保存するスパムラベル保存部と、前記スパムラベル保存部のURIスパムリストの問合せによって応答を行うスパムラベル応答部とを具備することを特徴とする。

【0012】

さらに、本発明において、前記スパムラベル管理サーバは、前記スパム処理サーバからURIスパムリストを取得するラベル取得部と、URIのリソースを取得して表示し当該リソースが有害か否かを確定するラベル確定部と、確定したURIを前記スパム処理サーバに送付して前記スパムラベル保存部の更新を指示するラベル保存部とを具備することを特徴とする。

【0013】

さらに、本発明において、前記フィルタリングエンジンは、前記URIスパムリストによってフィルタリングを行うか否かを含めてフィルタリング条件を設定するフィルタ設定管理部と、前記リソース閲覧用クライアントからのURI参照要求の際に前記スパム処理サーバからURIスパムリストを取得するスパムラベル取得部と、前記フィルタ設定管理部が設定したフィルタリング条件と前記スパムラベル取得部が取得したURIスパムリストとによってフィルタリングを行うフィルタリング処理部と、フィルタリング不要あるいは閲覧可能の場合に当該URIのリソースを取得しそれを前記リソース閲覧用クライアントに送付するリソース取得部とを具備することを特徴とする。

【0014】

さらに、本発明において、前記フィルタリングエンジンは、前記スパムラベル取得部が取得したURIスパムリストをキャッシュとして保存するスパムラベルキャッシュ部およびそれを制御するキャッシュ管理部を備え、前記リソース閲覧用クライアントからのURI参照要求の際に当該URIを既に取得しているか否か前記スパムラベルキャッシュ部をチェックし、取得済みの場合にはそれを参照することを特徴とする。

【0015】

さらに、本発明において、前記スパム処理サーバは、あらかじめキーワードを登録しているキーワード保存部と、前記解析処理部がスパムメールを解析する際に前記スパムメールに含まれるキーワードを抽出し、それを前記キーワード保存部のキーワードと比較して当該キーワードの有害度を判断し、前記スパムメールを格付けするスパムラベル格付部と

、前記スパムラベル格付部の格付結果に基づいて当該URIおよびそのスパムラベルをスパムラベル保存部に登録するスパムラベル登録部とを備えることを特徴とする。

【0016】

さらに、本発明のスパムメールフィルタリングシステムは、前記Webサイトにスパムメールを受信可能な複数のメールアドレスを有しそれぞれのメールアドレスに同一のスパムメールを受信し、前記複数のメールアドレスの一つが受信したスパムメールを受付けてそれを前記スパム処理サーバへ送付する第二スパム処理サーバを備え、前記スパム処理サーバでは自身が受付けたスパムメールと前記第二スパム処理サーバから送付されたスパムメールとが同一のスパムメールであることを認識して解析処理することを特徴とする。

【0017】

さらに、本発明において、前記スパム処理サーバは、前記複数のメールアドレスの一つが受信した第一のスパムメールと前記複数のメールアドレスの他の一つが受信した第二のスパムメールとを比較し両者が同一のスパムメールであることを認識する際、それぞれのメールの内容を突き合せて照合し同じ内容を含む照合率に基づいて判定することを特徴とする。

【0018】

また、本発明のスパムメールフィルタリング方法は、スパムメールを受け付け、前記スパムメール内に含まれるURIを収集しそれをURIスパムリストとして保持し、フィルタリングを行う際には、前記URIスパムリストを参照し、あらかじめ設定されているフィルタリング条件に基づいてフィルタリングすることを特徴とする。

【0019】

さらに、本発明の方法は、スパムメールを解析する際に前記スパムメールに含まれるキーワードを抽出し、それをあらかじめ登録してあるキーワードと比較して当該キーワードの有害度を判断し、それをスパムラベルとして前記スパムメールを格付けし、その格付結果に基づいて当該URIおよびそのスパムラベルを保持し、フィルタリングを行う際には、前記URIおよびそのスパムラベルを参照し、あらかじめ設定されているフィルタリング条件に基づいてフィルタリングすることを特徴とする。

【0020】

さらに、本発明の方法は、複数のメールアドレスの一つが受信した第一のスパムメールと前記複数のメールアドレスの他の一つが受信した第二のスパムメールとを比較し両者が同一のスパムメールであることを認識する際、それぞれのメールの内容を突き合せて照合し同じ内容を含む照合率に基づいて判定し、その判定結果を当該URIスパムリストに反映させることを特徴とする。

【0021】

また、本発明のスパムメールフィルタリングプログラムは、スパムメールを利用して情報のフィルタリングを行う際に、スパムメールを受け付けるスパム受付機能と、前記スパムメールの内容を解析する解析処理機能と、前記スパムメール内のURIをURIスパムリストとして登録するスパムリスト登録機能と、前記URIスパムリストを保存するスパムラベル保存機能と、前記URIスパムリストへの問合せによって応答を行うスパムラベル応答機能とを、コンピュータに実現させることを特徴とする。

【0022】

さらに、本発明のプログラムは、前記URIスパムリストによってフィルタリングを行うか否かを含めてフィルタリング条件を設定するフィルタ設定管理機能と、リソース閲覧用クライアントからのURI参照要求の際に前記URIスパムリストを取得するスパムラベル取得機能と、前記フィルタ設定管理機能が設定したフィルタリング条件と前記スパムラベル取得機能が取得したURIスパムリストとによってフィルタリングを行うフィルタリング処理機能と、フィルタリング不要あるいは閲覧可能な場合に当該URIのリソースを取得しそれを前記リソース閲覧用クライアントに送付するリソース取得機能とを、コンピュータに実現させることを特徴とする。

【0023】

さらに、本発明のプログラムは、スパムメールを解析する際に前記スパムメールに含まれるキーワードを抽出し、それをあらかじめ登録してあるキーワードと比較して当該キーワードの有害度を判断し、それをスパムラベルとして前記スパムメールを格付けし、その格付結果に基づいて当該URIおよびそのスパムラベルを保持し、フィルタリングを行う際には、前記URIおよびそのスパムラベルを参照し、あらかじめ設定されているフィルタリング条件に基づいてフィルタリングする機能を、コンピュータに実現させることを特徴とする。

【0024】

すなわち、本発明は、大規模電子ジャンクメール（以後、スパム）メーリングリストを利用した不特定多数への情報発信により、たとえば子供に望ましくない情報が掲載されているWebサイトの案内が配布され、閲覧されることを防止する。

【0025】

具体的には、インターネットのWebサイトにスパム受付用のダミーのメールアドレスを用意しておき、スパム発行者がこのメールアドレス宛てに送付したスパムメールに含まれるWebサイトのリソースを指すURIを保持し、モバイル環境やパソコンなどのWebブラウザ（リソース閲覧用クライアント）からこのスパムメールに含まれるURIのリソースを参照する場合に、保持したURIに基づいてフィルタリングする。

【発明の効果】

【0026】

以上、詳細に説明したように、本発明によれば、次の効果が得られる。

【0027】

第一の効果は、スパムメールが発行されるとすぐに当該メール内のURIはフィルタリングの対象となるため、即時性の効果があることである。従来は検索エンジンなどで検索したURIを格付けしていたため、たとえば1日だけ公開している違法な情報ページを含むURIのスパムメールについては、フィルタリングシステムは対応することができなかったことが、本発明により大きく改善される。

【0028】

第二の効果は、データの収集がスパムメールによる受動的、かつ自動の収集方法になることから、運用コストを大幅に削減できることがあげられる。従来のように、URIごとに格付けを行ったメタデータラベルを元にフィルタリングする手法では、そのラベル構築に人件費やサーバ運用費を含む膨大な費用がかかっていたが、本発明においてはサーバ運用費のみで実現可能であり、トータルコストを大幅に削減できる。

【0029】

第三の効果は、スパムメール発行者にとって、本発明が実現されることにより、スパムメール発行の価値がなくなり、スパムメールの発行そのものを抑制することができる。

【発明を実施するための最良の形態】

【0030】

以下、本発明について図面を参照しながら説明する。

【0031】

図1は本発明の実施の第一の形態を示す構成図である。同図において、本発明によるスパムメールフィルタリングシステムは、ネットワーク01と、Webページなどのリソースが存在するWebサイト02と、Webサイト02の中の任意のWebサイトに登録されたスパム受付用メールアドレス021と、このスパム受付用メールアドレス021を参照しスパムメールを発行するスパム発行端末03と、そのスパムメールを受け付けてURIスパムリスト151として自動的に保持するスパム処理サーバ10と、このスパムメールに含まれるURIを確認し誤りがあれば訂正するためのスパムラベル管理サーバ30と、リソース閲覧用クライアント04からのURIのリソース閲覧の際にスパム処理サーバ10からのURIスパムリストによって、あらかじめ設定した状態によって閲覧をフィルタリングするフィルタリングエンジン20とから構成されている。

【0032】

スパム処理サーバ10は、スパムメールを受けつけるスパム受付部11と、スパムメールの内容を解析する解析処理部12と、スパムメール内のURIをURIスパムリストとして登録するスパムリスト登録部13と、URIスパムリスト151を保存したスパムラベル保存部15と、このスパムラベル保存部15のURIスパムリスト151への問い合わせによって応答を行うスパムラベル応答部14とから構成されている。

【0033】

ここで、スパムラベル応答部14は問い合わせによってURIスパムリスト151そのものを返却してもよいし、存在有無のみを返却するものでもよい。また、URIスパムリスト151はURIのリストでもよいし、URIスパムリスト151を識別する加工値（ハッシュなど）であってもよい。

【0034】

また、解析処理部12が抽出するWebサイトのリソースを指すURIはスパムメールに含まれる1つある場合と複数ある場合、あるいはURIが存在しない場合もある。

【0035】

さらに、解析処理部12においては、当該メールがスパムメールか否かを自ら判断する処理を行うケースもある。具体的には、“_”文字で単語をつなげるパターン認識（This_is_a_spam_mail）や、有害な文字に空白を含めるケース（MakeMoney）のパターン認識により、スパムメールと判断しない場合は、これを破棄し、以後の処理を行わないオプションである。

【0036】

次に、スパムラベル管理サーバ30は、スパム処理サーバ10からURIスパムリスト151を取得するラベル取得部31と、URIのリソースを取得し、これを人が読める形で表示し、当該URIのリソースが有害か否か確認させ、その結果を確定するラベル確定部32と、確定したURIをスパム処理サーバ10のスパムラベル保存部15に保存するためのラベル保存部33とから構成されている。

【0037】

ここで、スパムラベル管理サーバ30はスパム処理サーバ10によるスパムメール受付と同時に動作してもよいし、後日、スパムラベル管理サーバ30の操作者によって動作してもよい。

【0038】

フィルタリングエンジン20は、あらかじめURIスパムリストによってフィルタリングを行うか否かを含めて設定が可能なフィルタ設定管理部21と、リソース閲覧用クライアント04からURI参照要求の際にスパム処理サーバ10からURIスパムリスト151を取得するためのスパムラベル取得部22と、URIスパムリスト151をキャッシュとして保存したスパムラベルキャッシュ部26と、それを管理するキャッシュ管理部25と、フィルタ設定管理部21が設定した状態およびスパムラベル取得部22が取得したURIスパムリストによってフィルタリングを行うフィルタリング処理部24と、フィルタリング不要の場合に当該URIのリソースを取得するリソース取得部23とから構成されている。

【0039】

ここで、スパムラベルキャッシュ部26はスパムラベル取得部22が取得したURIスパムリスト151を保存してもよいし、閲覧可能か否かという情報のみ、あるいはURIスパムリスト151を識別する加工値（ハッシュなど）であってもよい。

【0040】

図1において、スパム受付用メールアドレス021は、スパム処理サーバ10が受け付ける専用のメールアドレスであり、任意のWebサイト02に掲載しておく。スパム発行端末03はWebサイト02に掲載されたスパム受付用メールアドレス021を収集エンジンを使って収集し、収集したメールアドレスに対して大量のスパムメールを発行する。

【0041】

スパム処理サーバ10は、スパム受付部11において発行されたスパムメールを受信し

、その内容を解析処理部12において解析してスパムメール内に含まれるWebサイトのリソースを指すURIを抽出し、それをURIスパムリスト151としてスパムラベル保存部15にスパムリスト登録部13が登録を行う。

【0042】

上記の登録処理とは非同期に、スパムラベル管理サーバ30においては、スパム処理サーバ10が自動で保存したスパムラベル保存部15のURIスパムリスト151に誤りがないかを確認し、補正するために、ラベル取得部31がスパム処理サーバ10のスパムラベル応答部14を介してURIスパムリスト151を取得する。

【0043】

さらに、ラベル確定部32において、このURIが指すリソースを取得し、この内容が人が目視確認することにより、有害か否かの判断を行い、その結果によりラベル保存部33がスパム処理サーバ10のスパムリスト登録部13を介して、スパムラベル保存部15のURIスパムリスト151について、削除などの更新を行う。

【0044】

一方、リソース閲覧用クライアント04からフィルタリンクエンジン20を通してURIのリソースを閲覧する場合、あらかじめフィルタ設定管理部21によりスパムメールをフィルタリングする設定にしておく。フィルタリングエンジン20においては、リソース閲覧用クライアント04からの要求により参照しようとしているURIがスパムリストに登録されていないかを調べるため、スパムラベル取得部22がスパム処理サーバ10のスパムラベル応答部14からURIスパムリスト151を取得するか、あるいはURIスパムリスト151に登録されているかいないかという情報のみを取得する。

【0045】

続いて、フィルタリング処理部24によっては、当該URIがURIスパムリスト151に登録されていると判断した場合は、リソース閲覧用クライアント04に当該URIの閲覧はできない旨を返却し、あるいはURIスパムリスト151に登録されていないと判断した場合は、当該URIのリソースをリソース取得部23によって取得し、これをリソース閲覧用クライアント04に送付する。

【0046】

ここで、スパム処理サーバ10のスパムラベル応答部14から取得したURIスパムリスト151の情報はキャッシュ管理部25とスパムラベルキャッシュ部26により保存され、次のフィルタリング処理部24によるURI閲覧可否判断の際に、スパム処理サーバ10への問い合わせを行わないようキャッシュ処理することも可能である。

【0047】

図2は上記のスパムメールフィルタリングシステムにおけるスパムメール受付動作を示す流れ図である。

【0048】

まず、スパム発行端末03はWebサイト02にある不特定多数のメールアドレスを収集エンジンなどを使って収集し、収集したメールアドレスに対して大量のスパムメールを発行する。ここで、Webサイト02に登録されたスパム受付用メールアドレス021も収集され、スパム発行端末03はスパム受付用メールアドレス021にもスパムメールを発行する(F01)。

【0049】

スパム処理サーバ10ではスパム受付用メールアドレス021の受信をスパム受付部11が行っており、スパムメールの受信を行う(F02)。

【0050】

続いて、受信したスパムメールの内容を解析し、メール内に含まれるURIを抽出する(F03)。

【0051】

ここで、当該スパムメールにURIが含まれない場合は、そのまま当該スパムメールは破棄される。抽出したURIは、スパムリスト登録部13によってスパムラベル保存部1

5のURIスパムリスト151へ登録される(F04)。

【0052】

ここまでは自動的に処理されるものであるが、URIスパムリスト151に登録されたURIが、本当にフィルタリングすべき情報か否かを確認することもできるべきである。

【0053】

そのため、スパムラベル管理サーバ30においては、ラベル取得部31によってスパム処理サーバ10のURIスパムリスト151の要求を行い(F05)、スパム処理サーバ10のスパムラベル応答部14がURIスパムリスト151を返却し(F06)、このURIスパムリスト151から1つずつURIが指すWebサイトのリソースを参照し(F07)、これを人が読める形で表示し、当該URIのリソースが有害か否か目視確認させ、ラベル確定部32においてその結果を確定し(F08)、ラベル保存部33によってそのURIをスパム処理サーバ10に更新依頼する(F09)。

【0054】

スパム処理サーバ10はこの更新依頼によって、確定したURIスパムリスト151の更新を行う(F10)。ここで更新とは、当該URIをURIスパムリスト151から削除すること、あるいはURIスパムリスト151にある当該URIに無効フラグを付加するなどによって更新することを含む。

【0055】

図3は上記のスパムメールフィルタリングシステムにおけるフィルタリング動作を示す流れ図である。

【0056】

管理者(この場合、たとえば子供にスパムメールに含まれるWebリソースを閲覧させたくないと考える親である場合、あるいは社員が業務時間中に業務に関係ないスパムメールに含まれるWebリソースを閲覧させたくない上司など)は、まずフィルタリングエンジン20にスパムメールに含まれるURIをフィルタリングするための設定を、リソース閲覧用クライアント04からフィルタリングエンジン20に行っておく。あるいはリソース閲覧用クライアント04ではなく、フィルタリングエンジン20に管理者が直接設定してもよい(F21、F22)。

【0057】

設定の内容は、スパムメールに含まれるURI参照のフィルタリングを行うか否かという設定や、利用者に応じた設定(たとえば子供がリソース閲覧用クライアント04を使う場合フィルタするが、親が使う場合フィルタしないなど)でもよい。

【0058】

次に、情報閲覧者(この場合、たとえば子供や社員など)は、インターネットなどのネットワーク01上にあるWebサイト02のWebページや音楽、映像などWebリソースなどの情報であるリソースを、リソース閲覧用クライアント04を使い、フィルタリングエンジン20を介して、要求する(F23)。

【0059】

所望のURIのリソースを閲覧あるいは取得する場合、スパムラベル取得部22は、フィルタリングエンジン20の高速化のために、過去に使われた同じURIがURIスパムリストとしてスパムラベルキャッシュ部26に登録されていれば、キャッシュ管理部25を介して参照し(F24)、キャッシュに存在しない場合は、スパム処理サーバ10のスパムラベル応答部14からスパムラベル保存部15のURIスパムリスト151を入手する(F25、F26)。

【0060】

ここで、URIスパムリスト151はスパムメールに含まれていたURIか否かを判断するためのものであり、スパムメールに含まれていた実際のURIのリストでもよいし、URIスパムリスト151を識別する加工値(ハッシュなど)であってもよい。

【0061】

次に、フィルタリング処理部24は、当該URIがURIスパムリスト151に存在す

るか否かによってフィルタリングするか否かを判断し(F27)、フィルタリングすると判断した場合は、フィルタリングされたために閲覧できない旨のメッセージをリソース閲覧用クライアント04に返す(F28)。

【0062】

フィルタリングの必要がないと判断した場合は、URIが指すWebサイトのリソースをリソース取得部23によって取得し(F29)、これをリソース閲覧用クライアント04に返す(F30)。

【0063】

リソース閲覧用クライアント04は、フィルタリングエンジン20から返却されたフィルタリングされたために閲覧できない旨のメッセージか、あるいは当該URIのリソースを、情報閲覧者に提示する(F31)。

【0064】

図4は本発明の実施の第二の形態を示す構成図である。同図において、スパムメールフィルタリングシステムは、あらかじめ関連するキーワード171を登録しているキーワード保存部17と、スパムメールを解析する際にそれに含まれるキーワードもURIとともに抽出する解析処理部12aと、前記キーワードを登録されているものと比較しその有害度を格付けるスパムラベル格付部16と、この格付結果に基づいて当該URI152およびスパムラベル153をスパムラベル保存部15aに登録するスパムラベル登録部18とを備えるスパム処理サーバ10aを有している。

【0065】

なお、図4において、上記スパム処理サーバ10a以外の各構成ブロックは、図1において既述したスパムメールフィルタリングシステムと同一の機能構成を備えている。

【0066】

図5は、図4に示したスパムメールフィルタリングシステムのスパムメール受付動作を示す流れ図である。

【0067】

最初に、スパムメールが発行され(F41)、これを受信する(F42)までは同じであるが、スパム処理サーバ10aのスパム受付部11が受け付けたスパムメールを解析処理部12aにおいて解析する際に、URIのほかにスパムメールに含まれるキーワードまで抽出を行い、スパムラベル格付部16が、抽出したキーワードとあらかじめ登録されているキーワード保存部17のキーワード171を比較することにより、当該キーワードの有害度を判断する(F43、F44)。たとえば、有害なキーワードが含まれているほど、当該URIは有害であるとする。

【0068】

ここで、スパムメールからキーワードを抽出する方法は、スパムメールのヘッダ(発行者、あて先、経由など)に含まれる文字からキーワード抽出する場合や、スパムメール本体に含まれる本文からキーワードを抽出する場合のほか、スパムメール内のURIに含まれる文字列からキーワードを抽出する場合などがある。また、スパムラベル格付部16がキーワードを比較する方法は、文字列の完全一致や、あいまい検索(正規化表現)などを含む。

【0069】

次に、スパムラベル登録部18では、スパムラベル格付部16の格付結果により、当該URI(URI152)とその格付値(スパムラベル153)をスパムラベル保存部15aに登録する(F45)。

【0070】

ここまでは自動的に処理されるが、登録されたURIが、本当にフィルタリングすべき情報か否かを確認することもできるべきであることは、既述の通りである。

【0071】

すなわち、スパムラベル管理サーバ30においては、ラベル取得部31によってスパム処理サーバ10aにURI152およびスパムラベル153の要求を行い(F46)、ス

スパム処理サーバ10aのスパムラベル応答部14がそれらを返却する(F47)。

【0072】

このURI152をもとにURIが指すWebサイトのリソースを参照し(F48)、これを人が読める形で表示し、当該URIのリソースが有害か否か目視確認させ、ラベル確定部32においてその結果を確定し(F49)、確定したURIをラベル保存部33によってスパム処理サーバ10aに更新依頼を行う(F50)。

【0073】

スパム処理サーバ10aはこの更新依頼によって、確定したURIおよびスパムラベルにより更新を行う(F51)。ここで更新とは、当該URIをURIスパムリストから削除すること、あるいはURIスパムリストにある当該URIに無効フラグを付加するなどによって更新すること、あるいは当該URI152のスパムラベル153に入る格付値の変更を含む。

【0074】

なお、リソース閲覧時のフィルタリング動作は既述の通りである(図3参照)。

【0075】

図6は本発明の実施の第三の形態を示す構成図である。同図におけるスパムメールフィルタリングシステムは、スパム受信用のメールアドレスを専用に用意することができない場合で、複数の有効なメールアドレスを用意できる場合、スパムメールは同じ内容を複数のメールアドレスに配送する特徴があることを利用している。

【0076】

すなわち、図6において、Webサイト02には新たにスパムメール受付用メールアドレスB022が追加され、このアドレス宛のメールはスパム受付部B41を備えるスパム処理サーバB40で受け付けられた後、スパム処理サーバ10bに送付される。

【0077】

スパム処理サーバ10bは、スパム受付用メールアドレス021宛のメールを受け付けて解析すると共に、スパム処理サーバB40から送付されたメールも解析処理し、二重受付をチェックする解析処理部12bを備える。

【0078】

ここで、スパム受付用メールアドレス021とスパム受付用メールアドレスB022とは同じWebサイト02にあってもよし、異なるWebサイトにあっても、あるいは別のWebページにあってもよい。さらに、スパム受付用メールアドレスB022、スパム処理サーバB40は複数存在していてもよい。

【0079】

図7は、図6に示したスパムメールフィルタリングシステムのスパムメール受付動作を示す流れ図である。

【0080】

スパム発行端末03は、Webサイト02にある不特定多数のメールアドレスを収集エンジンなどを使って収集し、収集したメールアドレスに対して大量のスパムメールを発行する。その際、Webサイト02またはその他のサイトに登録されたスパム受付用メールアドレスB022も収集され、スパム発行端末03はスパム受付用メールアドレスB022にスパムメールを発行する(F61)。

【0081】

スパム処理サーバB40では、スパム受付用メールアドレスB022の受信をスパム受付部B41が行っており、スパムメールの受信を行い、受信したメールをスパム処理サーバ10bに渡す(F62)。

【0082】

さらに、スパム発行端末03は、同様のメールをスパム受付用メールアドレス021にも発行する(F63)。

【0083】

スパム処理サーバ10bでは、スパム受付用メールアドレス021の受信をスパム受付

部11が行っており、スパムメールの受信を行う(F64)。

【0084】

次に、解析処理部12によって、受信したスパムメールの内容を解析するが、まず、スパム処理サーバB40のスパム受付部B41から渡されたスパムメールと同期を取るため、2つの同じ本文のメールの付き合わせ確認を行い、本文が同じメールがスパム処理サーバ10bとスパム処理サーバB40の2カ所で受信したことを確認すれば、その何れかのスパムメールについて、メール内に含まれるURIを抽出する(F65)。

【0085】

ここで、当該スパムメールにURIが含まれない場合は、そのまま当該スパムメールは破棄される。

【0086】

また、2つの同じ本文のメールを把握する際にも、たとえば、メールアドレスの先頭部分のみを抽出したあて先を含むケースがあり(たとえば、ueda@webpage.or.jpであれば、ueda様へのご案内という本文を含むなど)、全く同じ本文か否かではなく、同じ内容を90%以上含む本文であるという照合率による比較が、必要となる。

【0087】

ここで、90%は例であり、設定は"@"文字を含むメールアドレス、あるいは「様へ」などのキーワードを用いることにより、メールの照合を行わない部分を指定するといった文字列指定も行うことが考えられる。

【0088】

また、スパム処理サーバ10bとスパム処理サーバB40が受信するスパムメールの順番はどちらのサーバが先でも、あるいは同時でもよく、それらのスパムメールを解析する解析処理部12bでは、複数のスパムメールを保存するキューを用いることが有効な手段でもある。

【0089】

さらに、スパム処理サーバB40に相当する機能を複数用意する場合は、それら全てのサーバがスパムメールを受信した場合に解析処理部12bを起動するか、あるいは2つ以上、もしくは1つだけでも解析処理部12bを起動するか特に規定しない。

【0090】

さて、抽出したURIはスパムリスト登録部13によってスパムラベル保存部15のURIスパムリスト151へ登録される(F66)。

【0091】

ここまで自動的に処理されるものであるが、URIスパムリスト151に登録されたURIが、本当にフィルタリングすべき情報か否かを確認することもできるべきである。

【0092】

すなわち、既述したように、スパムラベル管理サーバ30を任意に操作し、URIのリソースを目視確認することができる。図7において、F67～F72の処理は図2におけるF05～F10の処理と同じであるので、ここでは説明を繰返さない。

【0093】

なお、上記のスパムメールフィルタリングシステムにおいても、リソース閲覧時のフィルタリング動作は図3において説明した通りである。

【産業上の利用可能性】

【0094】

URIなどのデータ収集を機械的に行っており、それをフィルタリングに即時反映することができる。また、URIリソースの目視による確認チェックを行うことができるので、フィルタリングの品質を維持管理することが可能になる。したがって、情報管理部門にとって有用なフィルタリング手法を提供できる。

【図面の簡単な説明】

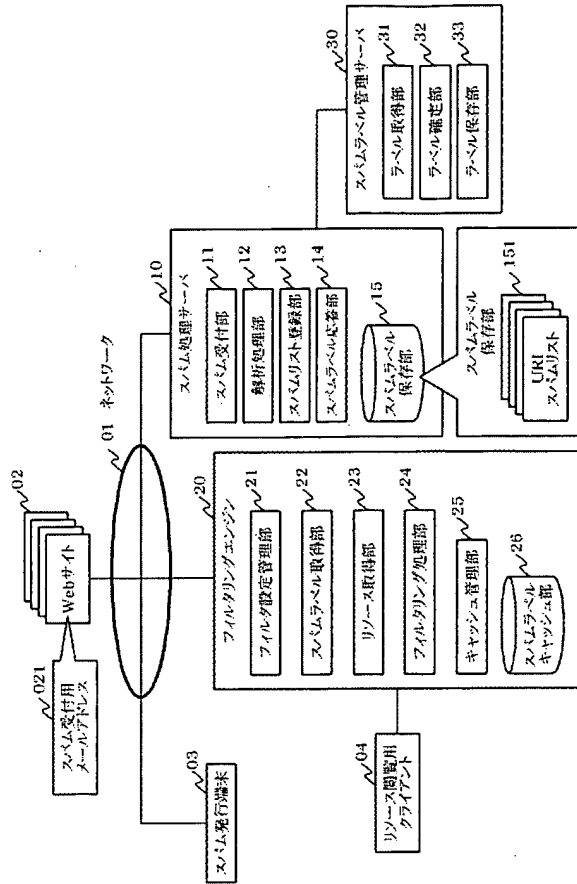
【0095】

- 【図1】本発明の実施の第一の形態を示す構成図。
【図2】スパムメールの受付動作を示す流れ図。
【図3】リソース閲覧時のフィルタリング動作を示す流れ図。
【図4】本発明の実施の第二の形態を示す構成図。
【図5】図4の実施例におけるスパムメール受付動作を示す流れ図。
【図6】本発明の実施の第三の形態を示す構成図。
【図7】図6の実施例におけるスパムメール受付動作を示す流れ図。
【符号の説明】

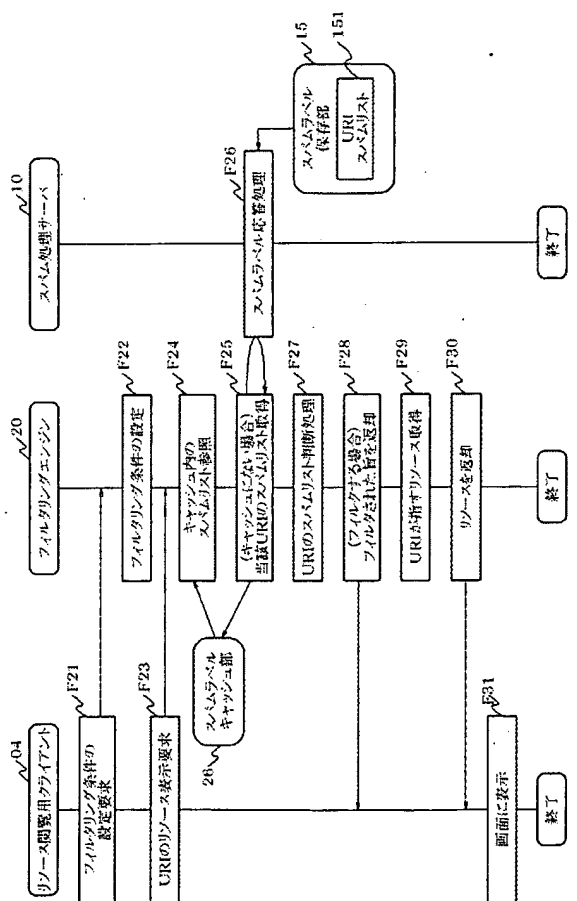
【0096】

- 01 ネットワーク
- 02 Webサイト
- 021 スпам受付用メールアドレス
- 022 スпам受付用メールアドレスB
- 03 スпам発行端末
- 04 リソース閲覧用クライアント
- 10, 10a, 10b スпам処理サーバ
- 11 スпам受付部
- 12, 12a, 12b 解析処理部
- 13 スпамリスト登録部
- 14 スпамラベル応答部
- 15, 15a スпамラベル保存部
- 16 スпамラベル格付部
- 17 キーワード保存部
- 18 スпамラベル登録部
- 20 フィルタリングエンジン
- 21 フィルタ設定管理部
- 22 スпамラベル取得部
- 23 リソース取得部
- 24 フィルタリング処理部
- 25 キャッシュ管理部
- 26 スпамラベルキャッシュ部
- 30 スпамラベル管理サーバ
- 31 ラベル取得部
- 32 ラベル確定部
- 33 ラベル保存部
- 40 スпам処理サーバB
- 41 スпам受付部B
- 151 URIスパムリスト
- 152 URI
- 153 スпамラベル
- 171 キーワード

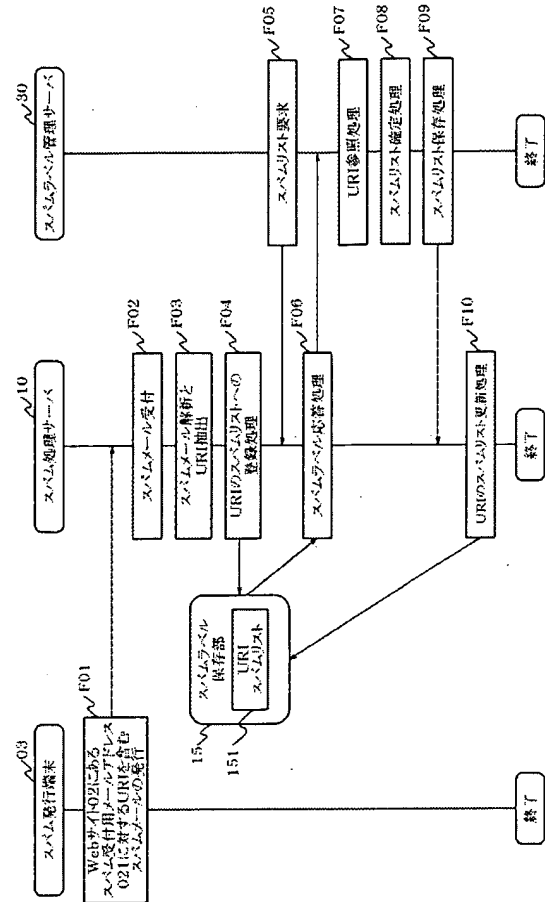
【図1】



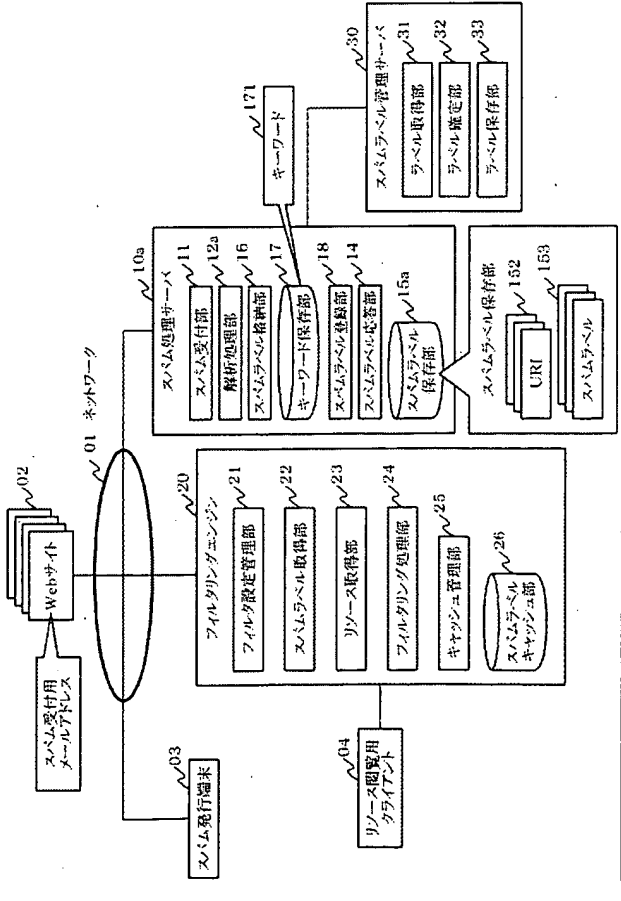
【図3】



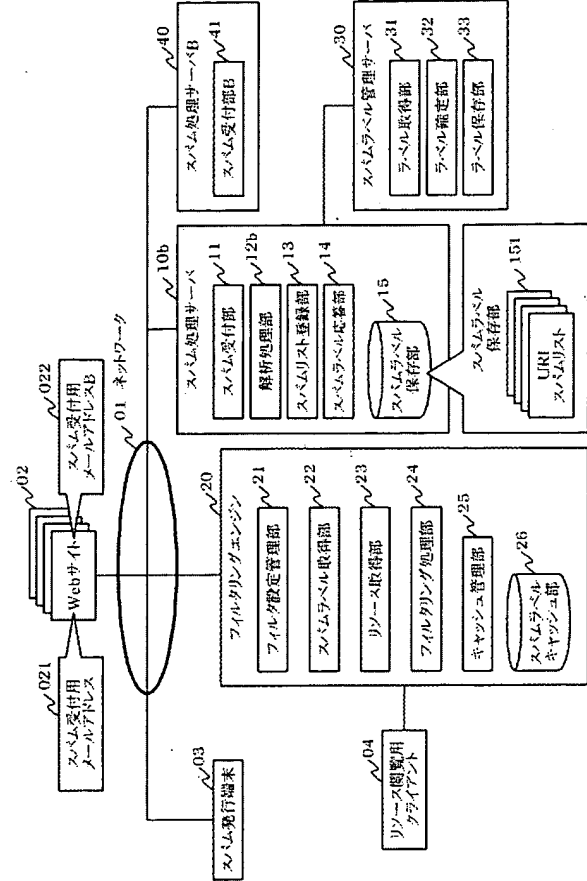
【図2】



【図4】



【図6】



【要約の続き】